Scientific Data Management Training

For PhD students specializing in Soil and Water Management hosted at

Sokoine University of Agriculture in Tanzania (SUA)

September 20-23, 2012
Imperial Botanical Entebbe, Kampala Uganda

Report by Resource Persons:

Margaret Nabasirye, Ph. D.

Department of Agricultural Production
College of Agricultural & Environmental Sciences
Makerere University
P.O. Box 7062, Kampala, Uganda

Tel: 256-77-519966
E-mail: mnabasirye@agric.mak.ac.ug

Susan Balaba Tumwebaze, Ph. D.

Department of Forestry, Biodiversity & Tourism
College of Agricultural & Environmental Sciences
Makerere University
P.O. Box 7062, Kampala, Uganda

Tel: 256-754-742065
E-mail: balaba2@yahoo.com

Executive Director, RUFORUM
Professor Adipala Ekwamu (e.adipala@ruforum.org, eadipala@agric.mak.ac.ug)

Sponsored by

RUFORUM

November 2012

1. INTRODUCTION

The issue of analyzing data to provide critical information required for policy formulation and publication of results in scientific data remains a major drawback in sub-Saharan Africa. It is upon every scientist in the region to participate in research activities that address food production constraints and socio-economic problems. This course aimed at providing young scientists involved in agricultural research systems with additional skills to manage, analyse, interpret and present results emanating from their data. The ultimate aim was to improve the efficient flow of agricultural information and research efficiency. This primary objective falls in line with the missions and goal of Regional Universities Forum for Capacity Building in Agriculture (RUFORUM) which aims at fostering "*innovativeness and adaptive capacity of universities engaged in agricultural and rural development to develop and sustain high quality in training, innovative and impact oriented research, and collaboration*".

The three-day course was held between 20 - 23 September, 2012 at the Imperial Botanical Beach Hotel, Entebbe, Uganda. This report gives an overview of the training course. The thrust of the training course was on scientific data management, covering aspects of data collection, data management, analysis, interpretation and reporting results. The main software used included Excel and Genstat computer packages. The course was practical, where participants obtained statistical experience through hands on statistics using their own data.

The training course was funded by Regional Universities Forum for Capacity Building in Agriculture (RUFORUM). RUFORUM sought and obtained support to organise the three day hands on statistics course for targeted postgraduate students involved in soil and water management research, sponsored by RUFORUM and hosted by SUA, who were at the stage of data analysis or thereabout. The training was held as a follow up of the scientific data management training held between 14 -18 November, 2011 at Bunda College of Agriculture from where the participants recommended follow-up training because at that time students from SUA had not collected data.

Previously, there has been a series of regional efforts to enhance capacity in research management in the region. For instance, recommendations of a meeting of 37 African Universities held in September 2009 in Mombasa, Kenya identified training in research skills as a priority need especially for young scientists and university lecturers. Additionally, RUFORUM organised a training of trainers (TOT) course (16-27 August 2010), aimed at building the capacity and/or identifying the needed human capacity in the region. The specific objective of the TOT was to ensure effective teaching and learning of statistics for research and information dissemination. The identified trainers from the sub-Saharan African region underwent training in Scientific Data Management course specifically on the content, preparation techniques, delivery methods and assessment approach. During the training, it was recommended that funding support be availed to train younger scientist in scientific data management throughout the region.

1.1 Training course objectives

The objective of this training was to update knowledge, skills and application capacity of handling research data, covering all aspects of collection, management, analysis, interpretation and reporting. To equip PhD students with the skills and knowledge in software use (GenStat and SPSS) in data management and analysis. The short term training would give them the hands-on skills they need to improve the quality and quantity of their research publications. The training was an opportunity for the postgraduate students to polish their draft publications. The specific objective was to empower students with knowledge and skills needed to:

1. Relate research objectives and hypotheses to data analysis.
2. Match different data types with appropriate methods of analysis.
3. Competently analyse research data, interpret, present and discuss results there from.
4. To improve skills in presentation of statistical results in various reports

In all aspects covered, participants were encouraged to relate these to their on-going research projects.

1.2 Approach

Participatory and interactive approaches were adopted. In general, short presentations were used to introduce or review the basic concepts of given topics and highlight their application. Facilitators closely interacted with participants to explore in-depth understanding and application of the concepts to research work. Participants got hands-on experience based on their research objectives, hypotheses/research questions and data generated from their own research studies. Hands-on practice with Excel, Genstat and SPSS resources occupied about three-quarters of each day's activities. Group work assignments for discussion and presentation was based on participants' work. Using their own data participants acquired an opportunity to analyse, interpret the results and prepare a draft data analysis report. Discussions on interpretation and presentation of the results were held everyday during the plenary sections. The last session of the course was allocated to presentations made by the participants on the analysis conducted on their data. The participants were then advised on the best ways to hand their data. This session reduced the fears of the participants in analysis their data

1.3 Course outline

The presentation by resources persons and related activities covered a range of topics (Appendix 1) as indicated in the table below:

Title of Session	Lead Resource Person(s)
Opening Session Registration Self introductions, expectations and agreement on content Software used– MS-Excel, Genstat/SPSS Participants' brief description of their studies and level in data collection	**SB**
Data management basics - Data and its importance, types of data, methods of data collection, record sheets, checking raw data, data entry and organization of files, integrity checks, outliers detection and remedy, data structure Practical using participants' data using Ms-Excel and Genstat	**SB**
Objectives, hypotheses and steps in data analysis - General steps in data analysis - Defining objectives of data analysis, w.r.t. study objectives/hypotheses/questions - Influence of study design and data type on types of analysis Practical on defining objectives of analysis, available data for these objectives	**MN**
Descriptive and exploratory data analysis - Organising and managing the data - Role of exploratory data analysis (EDA) - EDA methods/tools and when appropriate: tables, graphs, charts, pictures - Summary/descriptive statistics Practical on descriptive and EDA on participants' data using Excel and Genstat	**MN**
Confirmatory data analysis – inference and modeling - Concepts of inference and linkage with descriptive summary - Analysis of variance (anova) and statistical model - Assumptions and residual plots - Fixed, random and mixed effects models Genstat Practical on t-tests and ANOVA – General and REML - General linear model – correlation and regression - Assumptions and residual plots Genstat Practical on correlation and regression analysis	**SB**
Categorical data analysis - Categorical data, types (binary, multinomial) - Summary and descriptive table, cross tabulation, Chi-square tests - Generalized linear model Practical on Chi-square tests, generalized linear model and logistic model	**MN**
Applied multivariate data analysis - Principal component analysis (PCA), Factor analysis, cluster analysis (CA), canonical correlation analysis, discriminatory analysis Practical on multivariate analysis	**MN**
Writing up and presenting results - Effective means of presentation of results - Reporting on statistical methods used - Consideration of various audiences (farmers/scientists/policy) Participants' presentation of their results and short write-up.	**MN/SB**

Resource Persons: MN: Margaret Nabasirye; SB: Susan Balaba Tumwebaze

1.4 Attendance

The training targeted young scientists who are currently registered for their PhD in Soil and Water Management hosted at Sokoine University of Agriculture in Tanzania (SUA), but carrying out research at various locations in the region. A total of eight participants (Appendix 2) were trained against a target of 20, the rest did not turn up due to failure of some institutions to send representatives. The Participants had collected data and were in the process of conducting analysis. The training course was facilitated by two persons with a wide experience in research methodologies, data management and various statistical packages, namely: Dr. Margaret Nabasirye and Dr. Susan Balaba Tumwebaze.

.

1.5 Participants' Expectations and daily reviews

On the first day of the training course, self introductions were made; objectives and approach of the training were presented. The participants were requested to indicate their expectations. These expectations were synthesised and summarized as follows;

i. To gain practical skills and knowledge of scientific data management;
ii. To gain knowledge in multivariate analysis & interpretation;
iii. Confidentially use GENSTAT for data analysis and interpret of generated results
iv. Meet all objectives of course in an effective manner and cover as much as possible.

This information was used to re-structure the training programme in order to enhance participation and learning. The facilitators were flexible and willing to accommodate participants' suggestions and this contributed to the high level of attendance and individual participation.

1.5.1 Rules and guidelines

Participants agreed on the following group rules and guidelines: either to switch off mobile phones or keep them in a silent mode; respect for each other's views; and good time management. Consequently a time keeper was selected to ensure good time management and an energizer to keep the group awake and active.

1.6 Methods and tools

The basic concepts of each topic were introduced in form of PowerPoint presentations and followed by a demonstrated example using life data. At the end of every module, the participant analysed their data using Genstat Release 14 and facilitators guided them throughout the hands on practical. This approach enhanced their understanding of the modules taught; analytical skills; boosted confidence to handle their own data management and analysis. Presentations were backed up by power point handouts with various exercises, case

studies were used to reinforce learning and the flip charts used to illustrate issues. The participants brought their data and laptops to facilitate learning. The last session of the course was allocated to presentation made by the participants on the analysis conducted on their data. The participants were then advised on the best ways of handling their data. Practical examples and discussions illustrated application of computer based tools. The level of participation by all participants was high and the level of interaction between facilitators and students was very high.

1.7 Workshop evaluation

Daily evaluations were made on each day's activities covering aspects on module content &, delivery, depth of coverage, knowledge acquired, facilitators, time management and welfare. Results of overall evaluation indicated that;

i. The course was very important and useful and was conducted at the time when every participant has finished collecting data; the participants were at the point of data analysis and theses writing.
ii. Additionally participant felt that the time allocated to the training was not enough to practice.

Overall, the participants noted that the course was beneficial and it would be good to conduct such course twice in year, so that students/researcher are retooled in statistical analysis, interpretation and writing-up results.

2.0 OUTPUTS AND IMPACTS

This training course provided the SUA graduate students an opportunity to apply the knowledge and skills to their own research work. This training would minimize use of inappropriate statistical methods on data analysis and encourage publications. The overall evaluation of the workshop indicated that participants had learnt new concepts and gained useful skills applicable to their research set-ups and will be able to incorporate what they had learnt in future research work. The participants were committed and had a will to be trained and eager to learn and practice with Genstat.

2.1 Course learning outcomes

After the course the students are able to

- Competently handle different data types using Genstat statistical software
- Understand and apply correlation & regression analysis; multivariate techniques; categorical data analysis techniques; mixed (REML) models procedures & generalized linear models and also handle qualitative data.
- Use statistical software to describe, analyze and model the state of a biological or agricultural system in both a quantitative and qualitative manner.
- Make appropriate interpretation of results, presentation and reporting

With this training participants are now able to handle data analysis for their research and complete writing of dissertation within specific time frame; to communicate statistical findings effectively and to publish research findings in peer-reviewed journals and effectively disseminate these results for use by

stakeholders. Participants are also able to share ideas that enhance application of statistics in agricultural production systems;

3.0 CONCLUSIONS AND RECOMMENDATIONS

All participants indicated that the course was very relevant and that they had learnt a lot of applications during the training. Almost everyone found the exploratory and confirmatory data analysis aspects very useful as well as the data management.

The general impression is that participants were happy with the way the training proceeded. Their participation was fully explored by their involvement in several 'heated' discussions, for example the exciting graphs obtained during the multivariate analysis, etc. Interaction with facilitators and use of real examples based on their experience provided a conducive environment for learning. This was highlighted in their comments in the workshop evaluation. The participants felt that the time was short but the training was scheduled at the right time when all the participants had collected their own data. The training materials that were handed out at the beginning of the course, together with the CD of power point presentations were a useful resource for them.
The group was interested in the workshop activities and indeed hard working. Regarding participants, we had very good participants, committed, with a will to be trained and eager to learn and practice with the resources given out. We, the resource persons feel encouraged and challenged to organize/facilitate similar training workshops in the region.

4.0 ACKNOWLEGEMENTS

RUFORUM is sincerely acknowledged for funding this training, together with its role in initiating and facilitating the process. Special thanks go to Prof. Adipala Ekwamu, RUFORUM Executive secretary for all the logistical arrangements and Ms. Chindime Sylvia for effectively coordinating the training. We also thank the participants for the patience and active participation and commitment to learning. The onus is now on them to put the knowledge and skills in practice.

5.0 APPENDICES

Appendix 1: Time table
Appendix 2: List of participants

Appendix 1: Time table

Daily Training Programme: Scientific Data Management Course for PhD Students 20th -23/9/2012

Time	Friday 20/09/2012	**Saturday** 21/09/2012	**Sunday** 22/09/2012
0830-	**Opening Session** Registration Self introductions, expectations and agreement on content Software used– MS-Excel, Genstat/SPSS Participants' brief description of their studies and level in data collection **SB**	Confirmatory data analysis - inference and modeling/ Practical session **(SB)**	Applied multivariate data analysis/ Practical session **(MN/SB)**
1015-	***Health Break***	***Health Break***	***Health Break***
1030-	Data management basics)**/** Practical session**(SB**	Confirmatory data analysis/ Practical session **(SB)**	**Students presenting results (MN)**
1230-	***Lunch Break***	***Lunch Break***	***Lunch Break***
1330-	Objectives and steps in data analysis/ Practical session(**MN)**	Confirmatory data analysis/ Practical session **(SB)**	**Students presenting results (MN)**
1515-	***Health Break***	***Health Break***	***Health Break***
1530-	Descriptive and exploratory data analysis /Practical session**(MN)**	Categorical data analysis/ Practical session (**MN**)	Closing session Course evaluation **(SB**)

MN-Margaret Nabasirye; SB – Susan Balaba Tumwebaze;

Appendix 2: List of participants (in yellow are the nanes of the IDRC supported students)

	Name	Sex	Country
1.	Sixbert K. Mourice	Male	Malawi
2.	Anne Karuma	Female	Kenya
3.	Tolera Abera	Male	Ethiopia
4.	Msongalezi Barnabas	Male	Tanzania
5.	Singa, Darioin Dodonia	Male	Tanzania
6.	Sirikare Sylver	Male	Rwanda
7.	Mtandi John	Male	Malawi